

Article

The Distribution of Cross Sectional Momentum Returns When Underlying Asset Returns Are Student's t Distributed

Oh Kang Kwon ¹ and Stephen Satchell ^{2,*}

¹ Discipline of Finance, Codrington Building (H69), The University of Sydney, NSW 2006, Australia; ohkang.kwon@sydney.edu.au

² Trinity College, University of Cambridge, Cambridge CB2 1TQ, UK

* ses999gb@yahoo.co.uk

Received: 13 January 2020; Accepted: 31 January 2020; Published: 5 February 2020

Abstract: In Kwon and Satchell (2018), a theoretical framework was introduced to investigate the distributional properties of the cross-sectional momentum returns under the assumption that the vector of asset returns over the ranking and holding periods were multivariate normal. In this paper, the framework is extended to derive the corresponding results when the asset returns are multivariate Student's t . In particular, we derive the probability density function and the moments of the cross-sectional momentum returns and examine in detail the special case of two underlying assets to demonstrate that many of the salient features reported in the empirical literature are consistent with the theoretical implications.

Keywords: cross sectional momentum; student's t distribution; investment strategy

1. Introduction

Empirical investigation of the observed patterns in asset returns has been an active area of research in finance, with momentum, or persistence, in asset returns being one of the more popular examples of this line of research. Of these, perhaps the most prominent is cross-sectional momentum (CSM), which refers to the observation that the set of assets that outperform relative to another set over a prior period tend to continue to outperform over a subsequent period. The existence of CSM is usually tested empirically by sorting the assets according to their returns over a prior “ranking” period, and constructing a portfolio over a subsequent “holding” period by taking a long position in the “winners” and a short position in the “losers”. Statistically significant excess returns from following such a strategy would then support the existence of cross-sectional momentum.

Cross sectional momentum strategies are popular with practitioners since they tend to generate positive returns, while they are popular with academics due to the fact their existence would run contrary to an implication of the efficient market hypothesis that there does not exist any discernible patterns in asset returns. There is an extensive academic literature investigating the properties of CSM returns covering various asset classes, markets, and jurisdictions. The most notable findings are that CSM returns are generally slightly positive, but become highly negative during times of market uncertainty, and that losses during such periods tend to cancel out, or at least significantly reduce, the prior gains.

Various authors, including Fama and French (1992), Jegadeesh and Titman (1993 2001), Asness (1994), and Israel and Moskowitz (2013), found that momentum strategies are profitable in US equities markets over different time periods dating back to 1927. Analogous results were found for country equity indices by Richards (1997), Asness et al. (1997), Chan et al. (2000), and Hameed and Yuanto (2002), for emerging markets by Rouwenhorst (1998), for exchange rate markets in Okunev and White

(2003) and Menkhoff et al. (2012), for commodities by Erb and Harvey (2006), for futures contracts in Moskowitz et al. (2012), and in industries by Sefton and Scowcroft (2004). Similar results were also found by Asness et al. (2013) and Daniel and Moskowitz (2016) for markets in the European Union, Japan, the United Kingdom, and the United States, and across asset classes including fixed income, commodities, foreign exchange, and equity from 1972 through 2013.

Despite the extensive literature on the empirical properties of momentum based returns, there are relatively few that consider the distributional properties of these returns from a theoretical viewpoint, with Kwon and Satchell (2018) being a notable exception that addresses the CSM returns as defined in this paper. Most of the known theoretical results, obtained for example by Lo and MacKinlay (1990), Jegadeesh and Titman (1993), Lewellen (2002), and Moskowitz et al. (2012), are concerned only with the expected values and first order autocorrelations of returns from the so-called weighted relative strength strategy in which the portfolio over the holding period is constructed from all underlying assets weighted, essentially, in proportion to their absolute or relative returns over the ranking period. The reason why we wish to calculate the distribution of CSM returns is that we can then calculate percentiles, quantiles, and related quantities. We can deduce the degree to which moments of returns exist and their precise form. Such information can be used, for example, to assess the fatness of the tails of the distribution and this is valuable for risk management calculations as well as understanding the benefits and limitations of portfolio construction.

By assuming that underlying asset returns are Gaussian, the distribution and the moments of the CSM returns were derived in Kwon and Satchell (2018). In this paper, we extend their results to the case where the underlying asset returns are Student's t to derive the probability density function and the moments of the CSM returns. The t distribution arises naturally, for example, in a framework where asset volatility is stochastic, and conventional mean-variance analysis will create returns which are very similar to t -distributed returns. The important distinction between Student's t returns and normal returns is that the distribution of Student's t has an additional parameter which governs the fatness of the tails of the distribution and can be used to assess tail risk. There is a trade-off between realism and complexity; we would like to use a more complex distribution such as the skewed Student's t considered in Theodossiou (1998) and Hansen et al. (2010), but the analytical complexity that results becomes prohibitive.

Although the individual asset returns do not exhibit skewness under the generalization to Student's t , they can be leptokurtic which is a well-established feature in the empirical literature. Moreover, the CSM returns can, and do, exhibit skewness that depends on the statistical properties of the underlying assets. A detailed analysis of the special case of two underlying assets reveals that many of the salient features of the CSM returns reported in the empirical literature are consistent with the theoretical implications from this framework. This analysis is of interest because Kwon and Satchell (2018) were able to show that non-normality was a consequence of the momentum structure, even when the underlying returns were normal. We therefore wish to assess what the impact of assuming non-normality in the underlying returns will have on CSM returns. For example, will it exacerbate non-normality or make very little difference? Answers to this question will shed light on applying CSM to universes of assets which are fundamentally non-normal, such as emerging markets.

It should be pointed out that since we work under the assumption that asset returns over the ranking and holding periods are jointly t -distributed, there are limitations in the properties of momentum returns that can be addressed in the theoretical framework of this paper. For example, it is not possible to adequately address properties that depend on certain firm specific, economic, or financial factors such as liquidity, credit spread, market sentiment, business cycle, and information asymmetry since these factors cannot easily be captured in the distributional assumption on asset returns. Theoretical investigation of such properties would require an extension with the ability to incorporate such factors.

Finally, it may be asked what the connection is between our analysis and the extensive linear factor modelling that dominates the asset pricing literature. This literature essentially says that the time

t mean of an asset, say the first, is a linear function of factor returns. In the framework of this paper, we can accommodate such modelling by interpreting the asset mean to be conditional on factor returns.

The remainder of this paper is organized as follows: Section 2 introduces the notation and the key results on multivariate normal distributions, and Section 3 provides a mathematically precise definition of CSM returns. Although the expressions for the CSM return density and the associated moments are quite complex in general, they simplify considerably in the case of two assets with one winner and one loser, and this special case is examined in detail in Section 4, along with implications to the empirically observed features reported in the literature, and the paper concludes with Section 5.

2. Notation and Preliminaries

For the convenience of the reader, we introduce in this section the notation that will be used throughout the paper, and present some known results that will be relied upon in subsequent sections.

2.1. Notation

For any $x \in \mathbb{R}^n$, we will write x_i for the i -th coordinate of x , and given $y \in \mathbb{R}^n$ write $x \prec y$ if and only if $x_i < y_i$ for all $1 \leq i \leq n$. Similarly, given a matrix $M \in \mathbb{R}^{m \times k}$, we will write $M_{i,j}$ for the (i, j) -th entry of M , and the transpose of a vector or a matrix will be denoted by the superscript $'$. The vector in \mathbb{R}^n with all entries equal to 1 will be denoted $\mathbf{1}_n$, and given a subset $A \in \mathbb{R}^n$, we will denote by \mathbb{I}_A the indicator function on A .

Given a random vector, X , with values in a region $\mathcal{D}_X \subset \mathbb{R}^n$, we will write $f_X(x)$ and $F_X(x)$ for the probability density and the cumulative density functions of X , respectively. Moreover, given another random vector Y , with values in $\mathcal{D}_Y \subset \mathbb{R}^m$, we will denote by $f_{X|Y}(x|y)$ and $F_{X|Y}(x|y)$ the conditional probability density and conditional cumulative density functions of X given $Y = y$, respectively.

For any $n \in \mathbb{N}$, let $[n] = \{1, 2, \dots, n\}$ and let S_n be the set of permutations of $[n]$. We will denote the permutation that maps $1 \mapsto i_1, 2 \mapsto i_2, \dots, n \mapsto i_n$ by a sequence (i_1, i_2, \dots, i_n) , and given any $\tau \in S_n$ write $\tau(i)$ for the image of i under τ so that if $\tau = (1, 3, 2)$, for example, then $\tau(1) = 1$, $\tau(2) = 3$, and $\tau(3) = 2$. Given a permutation $\tau \in S_n$, we will denote by $P_\tau \in \mathbb{R}^{n \times n}$ the permutation matrix corresponding to τ and denote by $D_n \in \mathbb{R}^{(n-1) \times n}$ the matrix

$$D_n = \begin{bmatrix} -1 & 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & \cdots & 0 & 0 & 0 \\ & & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & -1 & 1 \end{bmatrix}. \quad (1)$$

The elements of the permutation group S_n act naturally on the set of polynomials, $\mathbb{R}[x_1, \dots, x_n]$ by the rule

$$\tau p(x_1, \dots, x_n) = p(x_{\tau_1}, \dots, x_{\tau_n})$$

for any polynomial $p \in \mathbb{R}[x_1, \dots, x_n]$ and $\tau \in S_n$. For any $n_1, n_2 \in \mathbb{N}$, let $p_{n_1, 2n_2} \in \mathbb{R}[x_1, \dots, x_{n_1+2n_2}]$ be the polynomial

$$p_{n_1, 2n_2}(x_1, \dots, x_{n_1+2n_2}) = \left(\prod_{i=1}^{n_1} x_i \right) \left(\prod_{i=1}^{n_2} (x_{n_1+2i} - x_{n_1+2i-1})^2 \right). \quad (2)$$

Denote by $Z(n_1, 2n_2)$ the stabilizer of $p_{n_1, 2n_2}$ under the action of $S_{n_1+2n_2}$ so that

$$Z(n_1, 2n_2) = \{\tau \in S_{n_1+2n_2} \mid \tau p_{n_1, 2n_2} = p_{n_1, 2n_2}\}, \quad (3)$$

and let $Q(n_1, 2n_2) = S_{n_1+2n_2}/Z(n_1, 2n_2)$ be the quotient group,¹ with elements of $Q(n_1, 2n_2)$ identified with their coset representatives $\tau \in S_{n_1+2n_2}$. Finally, define

$$[n]^m = \prod_{i=1}^m \{1, 2, \dots, n\} = \{(i_1, i_2, \dots, i_m) \mid 1 \leq i_j \leq n, 1 \leq j \leq m\} \quad (4)$$

as the m -fold Cartesian product of $[n] = \{1, 2, \dots, n\}$.

2.2. Multivariate Normal Distributions

The density of an n -dimensional normal distribution with mean μ and covariance Σ at $x \in \mathbb{R}^n$ will be denoted $\phi_n(x; \mu, \Sigma)$, and the corresponding cumulative density function will be denoted $\Phi_n(x; \mu, \Sigma)$. In general, given random variables X_1, \dots, X_n , their joint probability density function will be denoted f_{X_1, \dots, X_n} , and we will write F_{X_1, \dots, X_n} for the cumulative density function.

Theorem 1. Let $n_1, n_2 \in \mathbb{N}$ and suppose $X \sim \mathcal{N}_{n_1+n_2}(\mu, \Sigma)$, where

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}, \quad \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \Sigma_{1,1} & \Sigma_{1,2} \\ \Sigma_{2,1} & \Sigma_{2,2} \end{bmatrix}, \quad (5)$$

with $X_i, \mu_i \in \mathbb{R}^{n_i}$ and $\Sigma_{i,j} \in \mathbb{R}^{n_i \times n_j}$ for $1 \leq i, j \leq 2$, and Σ positive definite. Then, the conditional distribution of X_1 given X_2 is normal with mean and covariance

$$\mu_{X_1|X_2} = \mu_1 + \Sigma_{1,2}\Sigma_{2,2}^{-1}(X_2 - \mu_2), \quad (6)$$

$$\Sigma_{X_1|X_2} = \Sigma_{1,1} - \Sigma_{1,2}\Sigma_{2,2}^{-1}\Sigma_{2,1}, \quad (7)$$

respectively, and $\phi_{n_1+n_2}(x; \mu, \Sigma)$ decomposes as

$$\phi_{n_1+n_2}(x; \mu, \Sigma) = \phi_{n_1}(x_1; \mu_{X_1|X_2}, \Sigma_{X_1|X_2}) \phi_{n_2}(x_2; \mu_2, \Sigma_{2,2}). \quad (8)$$

Proof. Refer to Muirhead (1982) Theorem 1.2.11. \square

Given an n -dimensional random vector X and $p = (p_1, \dots, p_m) \in \{1, 2, \dots, n\}^m$, we will denote by $\mu_p(X)$ the p -th moment of X so that

$$\mu_p(X) = \mathbb{E} \left[\prod_{i=1}^m X_{p_i} \right]. \quad (9)$$

Note that the subscripts, p_i , in (9) may be repeated so that the above definition is equivalent to the more familiar definition of moments in which the powers of the components of X appear inside the expectation on the right-hand side, viz. $\mathbb{E} \left[\prod_{i=1}^n X_i^{k_i} \right]$ with $0 \leq k_i \in \mathbb{N}$ for $1 \leq i \leq n$. For example, we have $\mathbb{E} [X_1^2 X_2^3] = \mathbb{E} [X_1 X_1 X_2 X_2 X_2]$, and so in the notation of (9) we have $m = 5$, $d = (d_1, d_2, d_3, d_4, d_5) = (1, 1, 2, 2, 2)$, and $\mu_{(1,1,2,2,2)}(X) = \mathbb{E} [X_1^2 X_2^3]$.

¹ Refer to Rotman (1995) Chapter 2 for the details on quotient groups.

Theorem 2. Let $X \sim \mathcal{N}_n(\mu, \Sigma)$, where $n \in \mathbb{N}$, $\mu \in \mathbb{R}^n$ and $\Sigma \in \mathbb{R}^{n \times n}$ is positive definite. Then, for any $p = (p_1, \dots, p_m) \in \{1, 2, \dots, n\}^m$, we have

$$\mu_p(X) = \sum_{\substack{k, l \in \mathbb{N} \\ k+2l=m}} \sum_{\tau \in Q(k, 2l)} \left(\prod_{i=1}^k \mu_{p_{\tau_i}} \right) \left(\prod_{i=1}^l \Sigma_{p_{\tau_{k+2i-1}}, p_{\tau_{k+2i}}} \right). \quad (10)$$

In particular, if $\mu = \mathbf{0}$ and m is even, then

$$\mu_p(X) = \sum_{\tau \in Q(0, m)} \prod_{i=1}^{\frac{1}{2}m} \Sigma_{p_{\tau_{2i-1}}, p_{\tau_{2i}}}. \quad (11)$$

Proof. Refer to Withers (1985) Theorem 1.1. \square

An alternative expression for $\mu_p(X)$, where X is multivariate normal, is given in Kan (2008) Proposition 2.

Corollary 1. Let $X \sim \mathcal{N}_1(\mu, \sigma^2)$. Then, for any $m \in \mathbb{N}$, the m -th moment of X is given by

$$\mu_m(X) = \sum_{\substack{i=0 \\ i \text{ even}}}^m \binom{m}{i} (i-1)!! \sigma^i \mu^{m-i}, \quad (12)$$

where $k!! = \prod_{i=0}^{\lfloor \frac{k}{2} \rfloor} (k-2i)$ is the double factorial of $k \in \mathbb{N}$. In the special case where $\mu = 0$ and m is even, we have $\mu_m(X) = (m-1)!! \sigma^m$.

Proof. Follows from Theorem 2, since the inner sum in (10), for which $2l = i$, consists of $\binom{m}{i} (i-1)!!$ identical terms that are all equal to $\sigma^i \mu^{m-i}$. \square

2.3. Multivariate Student's t Distribution

Asset returns are often assumed to be normally distributed in the academic literature for theoretical convenience, in which case they are completely determined by the location and the scale parameters. However, it is widely reported in the empirical literature that the observed asset returns exhibit excess kurtosis. Student's t -distribution is a distribution from the elliptical family with an additional parameter ν , viz. number of degrees of freedom that controls the kurtosis. The Student's t distribution reduces to the normal in the limit as $\nu \rightarrow \infty$, and hence provides a convenient framework under which to investigate the impact of excess kurtosis in the underlying asset returns on the distributional properties of the CSM return. This subsection provides a brief summary of the key properties of the Student's t distribution that will be required in the remainder of this paper.

The t distribution has a long history in mathematical statistics. The univariate probability density function (pdf), $t(\mu, \sigma^2, \nu)$, of a t -variate with mean μ , scale parameter σ , and degrees of freedom ν is given by

$$t(\mu, \sigma^2, \nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sigma \sqrt{\nu \pi} \Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{1}{\nu} \left(\frac{x-\mu}{\sigma}\right)^2\right)^{-\frac{1}{2}(\nu+1)}. \quad (13)$$

From the origins of the t -test in mathematical finance, it is clear that we can write the corresponding random variable as $x = \mu + z/Y$, where z and Y are independent, $z \sim \mathcal{N}(0, \sigma^2)$, $Y = \sqrt{g/\nu}$, and $g \sim \chi^2(\nu)$.

In extending the definition of the t distribution to the multivariate case, we are faced with a choice. Although the choice $z \sim \mathcal{N}_n(\mathbf{0}, \Sigma)$ is clear, Y can be defined in various ways. For example:

- each component, z_i , of \mathbf{z} is normalized by independent $Y_i = \sqrt{g_i/v_i}$ with same or differing v_i ,
- Y_i could be jointly dependent,
- single common Y .

All three of these choices have a stochastic volatility interpretation corresponding to

- idiosyncratic shocks,
- common factor shocks,
- economy-wide or market-wide shock.

Although we have chosen the final characterization, cross sectional momentum could also be analyzed under other characterizations.

Throughout this paper, the probability density function of n -dimensional Student's t distribution, $\text{St}_{n_1+n_2}(\boldsymbol{\mu}, \Sigma, \nu)$, with ν degrees of freedom, location $\boldsymbol{\mu}$, and shape matrix Σ at $\mathbf{x} \in \mathbb{R}^n$ will be denoted $t_n(\mathbf{x}; \boldsymbol{\mu}, \Sigma, \nu)$, and we will write $T_n(\mathbf{x}; \boldsymbol{\mu}, \Sigma, \nu)$ for the corresponding cumulative density function. The next theorem shows that multivariate Student's t distribution is closed under conditioning, in the sense that the conditional density of a subset given its complement is again Student's t . This property will be crucial in the investigation of the CSM returns in later sections.

Theorem 3. Let $n_1, n_2 \in \mathbb{N}$ and suppose $\mathbf{X} \sim \text{St}_{n_1+n_2}(\boldsymbol{\mu}, \Sigma, \nu)$, where $\nu \in \mathbb{R}_+$,

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix}, \quad \boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \Sigma_{1,1} & \Sigma_{1,2} \\ \Sigma_{2,1} & \Sigma_{2,2} \end{bmatrix}, \quad (14)$$

with $\mathbf{X}_i, \boldsymbol{\mu}_i \in \mathbb{R}^{n_i}$ and $\Sigma_{i,j} \in \mathbb{R}^{n_i \times n_j}$ for $1 \leq i, j \leq 2$, and Σ positive definite. Then, the conditional distribution of \mathbf{X}_1 given \mathbf{X}_2 is Student's t with degrees of freedom $\nu_{\mathbf{X}_1|\mathbf{X}_2} = \nu + n_2$, and location and shape matrix

$$\boldsymbol{\mu}_{\mathbf{X}_1|\mathbf{X}_2} = \boldsymbol{\mu}_1 + \Sigma_{1,2}\Sigma_{2,2}^{-1}(\mathbf{X}_2 - \boldsymbol{\mu}_2), \quad (15)$$

$$\Sigma_{\mathbf{X}_1|\mathbf{X}_2} = \frac{\nu + (\mathbf{X}_2 - \boldsymbol{\mu}_2)' \Sigma_{2,2}^{-1} (\mathbf{X}_2 - \boldsymbol{\mu}_2)}{\nu + n_2} \left(\Sigma_{1,1} - \Sigma_{1,2}\Sigma_{2,2}^{-1}\Sigma_{2,1} \right), \quad (16)$$

respectively, and $t_{n_1+n_2}(\mathbf{x}; \boldsymbol{\mu}, \Sigma, \nu)$ decomposes as

$$t_{n_1+n_2}(\mathbf{x}; \boldsymbol{\mu}, \Sigma, \nu) = t_{n_1}(\mathbf{x}_1; \boldsymbol{\mu}_{\mathbf{X}_1|\mathbf{X}_2}, \Sigma_{\mathbf{X}_1|\mathbf{X}_2}, \nu_{\mathbf{X}_1|\mathbf{X}_2}) t_{n_2}(\mathbf{x}_2; \boldsymbol{\mu}_2, \Sigma_{2,2}, \nu). \quad (17)$$

Proof. Refer to Roth (2013) Appendix A.6, or Muirhead (1982) Problems 1.29 and 1.30. \square

Although a Student's t distribution does not have finite moments of all orders, the next theorem provides an explicit expression for those that do exist.

Theorem 4. Let $\mathbf{X} \sim \text{St}_n(\boldsymbol{\mu}, \Sigma, \nu)$, where $n \in \mathbb{N}$, $\nu \in \mathbb{R}_+$, $\boldsymbol{\mu} \in \mathbb{R}^n$ and $\Sigma \in \mathbb{R}^{n \times n}$ is positive definite. Moreover, for any $m \in \mathbb{N}$, denote by \mathcal{P}_m the set of subsets of the set $\{1, 2, \dots, m\}$ of size k such that $m - k$ is even, where $0 \leq k \leq m$. Then, for any $\mathbf{p} = (p_1, \dots, p_m) \in \{1, 2, \dots, n\}^m$ such that $m < \nu$, we have

$$\mu_{\mathbf{p}}(\mathbf{X}) = \sum_{P \in \mathcal{P}_m} \frac{\Gamma\left(\frac{\nu}{2} - \frac{1}{2}|P^c|\right)}{\Gamma\left(\frac{\nu}{2}\right)} \left(\frac{\nu}{2}\right)^{\frac{1}{2}|P^c|} \prod_{k \in P} \mu_{p_k} \sum_{\tau \in Q(0, |P^c|)} \left(\prod_{j=1}^{\frac{1}{2}|P^c|} \Sigma_{p_{i_{\tau_{2j-1}}}, p_{i_{\tau_{2j}}}} \right), \quad (18)$$

where $P^c = \{i_1, i_2, \dots, i_{k+2l}\}$ in the final sum.

Proof. Refer to Appendix A. \square

Setting $n = 1$ gives the moments of the one-dimensional Student's t distribution.

Corollary 2. Let $X \sim \text{St}_1(\mu, \sigma^2, \nu)$. Then, for any $m \in \mathbb{N}$ such that $m < \nu$, the m -th moment of X is given by

$$\mu_m(X) = \sum_{\substack{k=0 \\ k \text{ even}}}^m \binom{m}{k} (k-1)!! \frac{\Gamma\left(\frac{1}{2}(\nu-k)\right)}{\Gamma\left(\frac{\nu}{2}\right)} \left(\frac{\nu}{2}\right)^{\frac{k}{2}} \mu^{m-k} \sigma^k, \quad (19)$$

where $k!!$ is the double factorial defined in Corollary 1.

Proof. Follows from similar arguments to Theorem 4 noting that the indices p_{ij} are all equal in this case. \square

For an alternative derivation of the moments of Student's t distribution, refer to Kirkby et al. (2019).

2.4. Unified Skew t Family of Distributions

Multivariate skew-normal (SN) distributions were introduced in Azzalini and Valle (1985) to generalize normal distributions to those that allow non-zero skewness, and the seemingly disparate distributions related to the multivariate SN distributions were brought together under the umbrella of the so-called unified skew-normal (SUN) family of distributions in Arellano-Valle and Azzalini (2006), where it was shown that the SUN family contains many of these skew-normal variants as special cases. The extension of the normal family to those with non-zero skewness was then extended to the elliptical family of distributions in Arellano-Valle and Genton (2010). In what follows, we only summarize the results on the extension for the multivariate Student's t distributions that will be required in this paper, and refer the reader to Arellano-Valle and Genton (2010) and Jamalizadeh and Balakrishnan (2012) for the details.

Given $\nu \in \mathbb{R}_+$, $n_i \in \mathbb{N}$, $\mu_i \in \mathbb{R}^{n_i}$, $\Sigma_{i,j} \in \mathbb{R}^{n_1 \times n_2}$ for $1 \leq i \leq j$, where $\Sigma'_{2,1} = \Sigma_{1,2}$ and $\Sigma_{i,i}$ are positive definite, let

$$\begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} \sim \text{St}_{n_1+n_2} \left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{1,1} & \Sigma_{1,2} \\ \Sigma'_{1,2} & \Sigma_{2,2} \end{bmatrix}, \nu \right). \quad (20)$$

Then, the probability density function, $f_U(\mathbf{u})$, of an n_1 -dimensional unified skew t (SUT) distributed random variable, \mathbf{U} , associated with $(\mathbf{X}'_1, \mathbf{X}'_2)'$ given by

$$f_U(\mathbf{u}) = \frac{t_{n_2}(\mathbf{u}; \mu_2, \Sigma_{2,2}, \nu)}{T_{n_1}(\mathbf{0}; -\mu_1, \Sigma_{1,1}, \nu)} T_{n_1} \left(\mathbf{0}; -\mu_1 - \Sigma_{1,2} \Sigma_{2,2}^{-1} (\mathbf{u} - \mu_2), \Sigma_{1,1,2}(\mathbf{u}), \nu + n_2 \right), \quad (21)$$

where

$$\Sigma_{1,1,2}(\mathbf{u}) = \frac{\nu + (\mathbf{u} - \mu_2)' \Sigma_{2,2}^{-1} (\mathbf{u} - \mu_2)}{\nu + n_2} \left(\Sigma_{1,1} - \Sigma_{1,2} \Sigma_{2,2}^{-1} \Sigma_{2,1} \right). \quad (22)$$

The key characteristic of $f_U(\mathbf{u})$ is that it is a product of an n_2 -dimensional Student's t density and an n_1 -dimensional cumulative Student's t density with the variable \mathbf{u} appearing as the main variable in the former and in the mean and variance parameters of the latter. As will be seen, the densities of cross sectional momentum returns will be a weighted sum of these SUT distributions.

3. Cross-Sectional Momentum Returns with Student's t Distributed Asset Returns

In this section, we derive the distributional properties of the cross sectional momentum (CSM) returns under the assumption that the underlying asset returns are multivariate Student's t . We begin by recalling the mathematically precise definition of the CSM return from Kwon and Satchell (2018).

Let $0 < m_+, m_-, n \in \mathbb{N}$ such that $m_+ + m_- \leq n$, and for each $1 \leq i \leq n$ denote by $r_{i,t}$ the return on asset i at time t . Moreover, let

$$\mathbf{r}_t = (r_{1,t}, \dots, r_{n,t})' \in \mathbb{R}^n, \quad (23)$$

and for any $\tau \in S_n$, define

$$r_{\tau, m_{\pm}, t} = \frac{1}{m_+} \sum_{i=1}^{m_+} r_{\tau_i, t} - \frac{1}{m_-} \sum_{i=1}^{m_-} r_{\tau_{n-m_-+i}, t} \in \mathbb{R}, \quad (24)$$

$$x_{\tau_i, t} = r_{\tau_{i+1}, t} - r_{\tau_i, t} \in \mathbb{R}, \quad 1 \leq i \leq n-1, \quad (25)$$

$$\mathbf{x}_{\tau, t} = (x_{\tau_1, t}, \dots, x_{\tau_{n-1}, t})' \in \mathbb{R}^{n-1}, \quad (26)$$

$$\mathbf{z}_{\tau, t} = (r_{\tau, m_{\pm}, t+1}, \mathbf{x}'_{\tau, t})' \in \mathbb{R}^n. \quad (27)$$

Note that any given $\tau \in S_n$ defines an ordering, $r_{t, \tau_1} > r_{t, \tau_2} > \dots > r_{t, \tau_n}$, of the components of \mathbf{r}_t . Thus, $r_{\tau, m_{\pm}, t}$ represents the return on a portfolio where the top m_+ ranked assets are equally weighted and held long while the bottom m_- assets are equally weighted and held short. The assumption of equal weighting is for notational simplicity only, and not crucial for the general theoretical results. Note also that $\mathbf{x}_{\tau, t}$ is defined to allow the ranking of the components of \mathbf{r}_t corresponding to $\tau \in S_n$ to be written succinctly as $\mathbf{x}_{\tau, t} \prec \mathbf{0}_{n-1}$.

Definition 1. The (m_+, m_-) -cross sectional momentum return, $r_{m_{\pm}, t+1}$, is defined by

$$r_{m_{\pm}, t+1} = \sum_{\tau \in S_n} \mathbb{I}_{\{\mathbf{x}_{\tau, t} \prec \mathbf{0}_{n-1}\}} r_{\tau, m_{\pm}, t+1}, \quad (28)$$

where \mathbb{I}_A , for any subset $A \subset \mathbb{R}^n$, denotes the indicator function on the set A .

For intuition behind the definition of $r_{m_{\pm}, t+1}$, note that the components of \mathbf{r}_t , representing asset returns over the ranking period, can be arranged in any of the $n!$ orderings corresponding to the permutations $\tau \in S_n$. For each such ranking $r_{t, \tau_1} > r_{t, \tau_2} > \dots > r_{t, \tau_n}$, the m_+ winner returns over the holding period are $r_{t+1, \tau_1}, \dots, r_{t+1, \tau_{m_+}}$ while the m_- loser returns are $r_{t+1, \tau_{n-m_-+1}}, \dots, r_{t+1, \tau_n}$. Equally weighting the returns in the winner and the loser portfolios gives $r_{\tau, m_{\pm}, t+1}$, and since the ranking of components of \mathbf{r}_t determined by $\tau \in S_n$ is equivalent to the condition $\mathbf{x}_{\tau, t} \prec \mathbf{0}_{n-1}$, summing over all possible $r_{\tau, m_{\pm}, t+1}$ and prefixing by the matching indicator function gives the expression for $r_{m_{\pm}, t+1}$ in (28).

For the remainder of this paper, we make the following assumption on the distribution of $(\mathbf{r}'_t, \mathbf{r}'_{t+1})'$.

Assumption 1. The vector of returns, $(\mathbf{r}'_t, \mathbf{r}'_{t+1})'$, is multivariate Student's t distributed so that

$$\begin{bmatrix} \mathbf{r}_{t+1} \\ \mathbf{r}_t \end{bmatrix} \sim \text{St}_{2n} \left(\begin{bmatrix} \boldsymbol{\mu}_{t+1} \\ \boldsymbol{\mu}_t \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{t+1, t+1} & \boldsymbol{\Sigma}_{t+1, t} \\ \boldsymbol{\Sigma}_{t, t+1} & \boldsymbol{\Sigma}_{t, t} \end{bmatrix}, \nu \right), \quad (29)$$

with $\nu \in \mathbb{R}_+$, $\boldsymbol{\mu}_u \in \mathbb{R}^n$, and $\boldsymbol{\Sigma}_{u, v} \in \mathbb{R}^{n \times n}$, where $u, v \in \{t, t+1\}$.

Since the t -distribution is symmetric, there are limitations on the properties of asset returns that can be captured adequately by the above assumption as already discussed in Section 1. Nevertheless, the assumption is sufficiently general to accommodate linear factor models and econometric models such as vector autoregressive moving average models where the factors and noise terms, respectively, are t -distributed. Moreover, the framework also allows consideration of more general cases where $\boldsymbol{\mu}_t$ and $\boldsymbol{\mu}_{t+1}$ are conditional means linear in factors without requiring the factors themselves to be

multivariate t . However, since the analysis in later sections will show that momentum returns are nonlinear in the underlying asset returns, the common practise of regressing momentum returns on the various factors must be interpreted as a best linear prediction rather than a conditional expectation in such cases. We now derive the probability density function, $f_{m_{\pm},t+1}(r)$, of the CSM return $r_{m_{\pm},t+1}$.

Theorem 5. Suppose that $(\mathbf{r}'_t, \mathbf{r}'_{t+1})'$ satisfies Assumption 1. Then, the probability density function, $f_{m_{\pm},t+1}(r)$, of the cross sectional momentum return, $r_{m_{\pm},t+1}$, is given by

$$f_{m_{\pm},t+1}(r) = \sum_{\tau \in S_n} \mathbf{t}_1(r; \mathbf{r}'_{\tau} P_{\tau} \boldsymbol{\mu}_{t+1}, \mathbf{r}'_{\tau} P_{\tau} \Sigma_{t+1,t+1} P'_{\tau} \boldsymbol{\nu}, \nu) T_{n-1}(\mathbf{0}; \lambda_{\tau}(r), \Lambda_{\tau}(r), \nu + 1), \quad (30)$$

where $\boldsymbol{\nu} = (\mathbf{1}'_{m_+}, \mathbf{0}'_{n-m_+-m_-}, -\mathbf{1}'_{n-})'$,

$$\lambda_{\tau}(r) = D_n P_{\tau} \boldsymbol{\mu}_t + \frac{(r - \mathbf{r}'_{\tau} P_{\tau} \boldsymbol{\mu}_{t+1})}{\mathbf{r}'_{\tau} P_{\tau} \Sigma_{t+1,t+1} P'_{\tau} \boldsymbol{\nu}} D_n P_{\tau} \Sigma_{t,t+1} P'_{\tau} \boldsymbol{\nu}, \quad (31)$$

$$\Lambda_{\tau}(r) = \frac{\nu + (\mathbf{r}'_{\tau} P_{\tau} \Sigma_{t+1,t+1} P'_{\tau} \boldsymbol{\nu})^{-1} (r - \mathbf{r}'_{\tau} P_{\tau} \boldsymbol{\mu}_{t+1})^2}{\nu + 1} \left(D_n P_{\tau} \Sigma_{t,t} P'_{\tau} D'_n - \frac{D_n P_{\tau} \Sigma_{t,t+1} P'_{\tau} \boldsymbol{\nu} \mathbf{r}'_{\tau} P_{\tau} \Sigma_{t+1,t} P'_{\tau} D'_n}{\mathbf{r}'_{\tau} P_{\tau} \Sigma_{t+1,t+1} P'_{\tau} \boldsymbol{\nu}} \right), \quad (32)$$

and we define $T_0 \equiv 1$. Alternatively,

$$f_{m_{\pm},t+1}(r) = \sum_{\tau \in S_n} \int_{\mathbb{R}^{n-1}} \mathbf{t}_1(r; \gamma_{\tau}(\mathbf{x}), Y_{\tau}(\mathbf{x}), \nu + n - 1) \mathbf{t}_{n-1}(\mathbf{x}; D_n P_{\tau} \boldsymbol{\mu}_t, D_n P_{\tau} \Sigma_{t,t} P'_{\tau} D'_n, \nu) d\mathbf{x}, \quad (33)$$

where

$$\gamma_{\tau}(\mathbf{x}) = \mathbf{r}'_{\tau} P_{\tau} \boldsymbol{\mu}_{t+1} + \mathbf{r}'_{\tau} P_{\tau} \Sigma_{t+1,t} P'_{\tau} D'_n (D_n P_{\tau} \Sigma_{t,t} P'_{\tau} D'_n)^{-1} (\mathbf{x} - D_n P_{\tau} \boldsymbol{\mu}_t), \quad (34)$$

$$Y_{\tau}(\mathbf{x}) = \frac{\nu + (\mathbf{x} - D_n P_{\tau} \boldsymbol{\mu}_t)' (D_n P_{\tau} \Sigma_{t,t} P'_{\tau} D'_n)^{-1} (\mathbf{x} - D_n P_{\tau} \boldsymbol{\mu}_t)}{\nu + n - 1} \left(\mathbf{r}'_{\tau} P_{\tau} \Sigma_{t+1,t+1} P'_{\tau} \boldsymbol{\nu} - \mathbf{r}'_{\tau} P_{\tau} \Sigma_{t+1,t} P'_{\tau} D'_n (D_n P_{\tau} \Sigma_{t,t} P'_{\tau} D'_n)^{-1} D_n P_{\tau} \Sigma_{t,t+1} P'_{\tau} \boldsymbol{\nu} \right). \quad (35)$$

Proof. Refer to Appendix B. \square

Note that the summands that appear in the pdf of the CSM return in (30) have the characteristic form of the SUT densities given in (21) other than for the omission of the normalization factor² that appears in the denominator of (21). It follows that pdf of the CSM return is a weighted sum of the SUT densities. The next result gives the pdf in the special case where \mathbf{r}_t and \mathbf{r}_{t+1} are independent, which can be considered as the case where the market is efficient.

Corollary 3. If $(\mathbf{r}'_t, \mathbf{r}'_{t+1})'$ satisfies Assumption 1, and \mathbf{r}_t and \mathbf{r}_{t+1} are independent, then

$$f_{m_{\pm},t+1}(r) = \sum_{\tau \in S_n} \mathbf{t}_1(r; \mathbf{r}'_{\tau} P_{\tau} \boldsymbol{\mu}_{t+1}, \mathbf{r}'_{\tau} P_{\tau} \Sigma_{t+1,t+1} P'_{\tau} \boldsymbol{\nu}, \nu) T_{n-1}(\mathbf{0}; D_n P_{\tau} \boldsymbol{\mu}_t, \Lambda_{\tau}^{\circ}(r), \nu + 1), \quad (36)$$

where

$$\Lambda_{\tau}^{\circ}(r) = \frac{\nu + (\mathbf{r}'_{\tau} P_{\tau} \Sigma_{t+1,t+1} P'_{\tau} \boldsymbol{\nu})^{-1} (r - \mathbf{r}'_{\tau} P_{\tau} \boldsymbol{\mu}_{t+1})^2}{\nu + 1} D_n P_{\tau} \Sigma_{t,t} P'_{\tau} D'_n, \quad (37)$$

² The normalization factor is, in fact, the probability of the event $\mathbf{0} \prec \mathbf{X}_1$, where \mathbf{X}_1 is as defined in (20).

Proof. Follows from Theorem 5 since $\Sigma_{t,t+1} = O_{n \times n} = \Sigma_{t+1,t}$ in this case. \square

We next derive the expressions for the non-central moments of the CSM returns. Since t distributions do not have moments of all orders as noted in Theorem 4, the moments of CSM returns will also only exist up to a certain order.

Theorem 6. Suppose $(\mathbf{r}'_t, \mathbf{r}'_{t+1})'$ satisfies Assumption 1, and let $m \in \mathbb{N}$ such that $m < \nu$. Then, the m -th non-central moment of $r_{m\pm, t+1}$ is given by

$$\mu_m(r_{m\pm, t+1}) = \sum_{\tau \in S_n} \sum_{\substack{k=0 \\ k \text{ even}}}^m \binom{m}{k} (k-1)!! \frac{\Gamma\left(\frac{1}{2}(\nu+n-1-k)\right)}{\Gamma\left(\frac{1}{2}(\nu+n-1)\right)} \left(\frac{\nu+n-1}{2}\right)^{\frac{k}{2}} \int_{\mathbb{R}^{n-1}} \gamma_\tau^{m-k}(\mathbf{x}) Y_\tau^{\frac{k}{2}}(\mathbf{x}) \mathbf{t}_{n-1}(\mathbf{x}; D_n P_\tau \boldsymbol{\mu}_t, D_n P_\tau \Sigma_{t,t} P_\tau' D_n', \nu) d\mathbf{x}, \quad (38)$$

where $\gamma_\tau(\mathbf{x})$ and $Y_\tau(\mathbf{x})$ are as defined in (34) and (35), respectively.

Proof. Refer to Appendix C. \square

4. Special Case of Two Assets

In this section, we examine in detail the special case of two assets, and begin by computing the partial moments of one-dimensional Student's t distributions that will be required. To reduce notational burden, we define for $\eta \in \mathbb{R}$, $\varsigma \in \mathbb{R}_+$, and $\nu \in \mathbb{R}_+$

$$c(\eta, \varsigma^2, \nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\varsigma \sqrt{\nu\pi} \Gamma\left(\frac{\nu}{2}\right)}, \quad (39)$$

so that from (13) we have

$$\mathbf{t}_1(x; \eta, \varsigma^2, \nu) = c(\eta, \varsigma^2, \nu) \left(1 + \frac{1}{\nu} \left(\frac{x-\eta}{\varsigma}\right)^2\right)^{-\frac{1}{2}(\nu+1)}. \quad (40)$$

Lemma 1. Let $\eta \in \mathbb{R}$, $\varsigma \in \mathbb{R}_+$ and $2 < \nu \in \mathbb{R}_+$. Then, for $m \in \mathbb{N}_+$, we have

$$\left(\frac{x-\eta}{\varsigma}\right)^m \mathbf{t}_1(x; \eta, \varsigma^2, \nu) = -\varsigma \sqrt{\frac{\nu}{\nu-2}} \left(\frac{x-\eta}{\varsigma}\right)^{m-1} \frac{\partial \mathbf{t}_1(x; \eta, \varsigma^2 \nu / (\nu-2), \nu-2)}{\partial x}. \quad (41)$$

Proof. Refer to Appendix D. \square

The next theorem will play a key role in the derivation of the non-central moments of the CSM returns.

Theorem 7. For any $\eta \in \mathbb{R}$, $\varsigma \in \mathbb{R}_+$, $\nu \in \mathbb{R}_+$, and $m \in \mathbb{N}$ such that $m < \nu$, let

$$\kappa_m(\eta, \varsigma^2, \nu) = \int_{-\infty}^0 \left(\frac{x-\eta}{\varsigma}\right)^m \mathbf{t}_1(x; \eta, \varsigma^2, \nu) dx. \quad (42)$$

Then, $\kappa_0(\eta, \varsigma^2, \nu) = \mathbf{T}_1(0; \eta, \varsigma^2, \nu)$, $\kappa_1(\eta, \varsigma^2, \nu) = -\varsigma \sqrt{\nu/(\nu-2)} \mathbf{t}_1(0; \eta, \varsigma^2 \nu / (\nu-2), \nu-2)$, and

$$\kappa_m(\eta, \varsigma^2, \nu) = (m-1) \left(\frac{\nu}{\nu-2}\right)^{\frac{1}{2}(m-1)} \kappa_{m-2}\left(\eta, \frac{\varsigma^2 \nu}{\nu-2}, \nu-2\right) \quad (43)$$

for $2 \leq m < v$. More explicitly, if m is even

$$\kappa_m(\eta, \zeta^2, v) = \frac{(m-1)!! v^{\frac{1}{2}(m-1)}}{(v-2)(v-4) \cdots (v-m+2) \sqrt{v-m}} T_1\left(0; \eta, \frac{\zeta^2 v}{v-m}, v-m\right), \quad (44)$$

and if m is odd

$$\begin{aligned} \kappa_m(\eta, \zeta^2, v) \\ = \frac{-\zeta(m-1)!! v^{\frac{1}{2}(m-1)}}{(v-2)(v-4) \cdots (v-(m-1)) \sqrt{v-(m+1)}} t_1\left(0; \eta, \frac{\zeta^2 v}{v-(m+1)}, v-(m+1)\right). \end{aligned} \quad (45)$$

Proof. Firstly, the expression for $\kappa_0(\eta, \zeta^2, v)$ follows directly from the definition of $\kappa_m(\eta, \zeta^2, v)$, and, for $\kappa_1(\eta, \zeta^2, v)$, we obtain on setting $m = 1$ in Lemma 1 that

$$\begin{aligned} \kappa_1(\eta, \zeta^2, v) &= -\zeta \sqrt{\frac{v}{v-2}} \int_{-\infty}^0 \frac{\partial t_1(x; \eta, \zeta^2 v / (v-2), v-2)}{\partial x} dx \\ &= -\zeta \sqrt{\frac{v}{v-2}} t_1\left(0; \eta, \frac{\zeta^2 v}{v-2}, v-2\right). \end{aligned}$$

Next, for (43), using Lemma 1 and applying integration by parts gives

$$\begin{aligned} \kappa_m(\eta, \zeta^2, v) &= -\zeta \sqrt{\frac{v}{v-2}} \int_{x \in \mathbb{R}_-} \left(\frac{x-\eta}{\zeta}\right)^{m-1} \frac{\partial t_1(x; \eta, \zeta^2 v / (v-2), v-2)}{\partial x} dx \\ &= (m-1) \sqrt{\frac{v}{v-2}} \int_{x \in \mathbb{R}_-} \left(\frac{x-\eta}{\zeta}\right)^{m-2} t_1\left(x; \eta, \frac{\zeta^2 v}{v-2}, v-2\right) dx \\ &= (m-1) \left(\frac{v}{v-2}\right)^{\frac{1}{2}(m-1)} \int_{x \in \mathbb{R}_-} \left(\frac{x-\eta}{\zeta \sqrt{v/(v-2)}}\right)^{m-2} t_1\left(x; \eta, \frac{\zeta^2 v}{v-2}, v-2\right) dx \\ &= (m-1) \left(\frac{v}{v-2}\right)^{\frac{1}{2}(m-1)} \kappa_{m-2}\left(\eta, \frac{\zeta^2 v}{v-2}, v-2\right), \end{aligned}$$

which is (43). The explicit expressions for the odd and even cases follow by induction. \square

As it will be seen, the quantities that play a key role in the two asset case are the spreads, $r_{t,2} - r_{t,1}$ and $r_{t+1,2} - r_{t+1,1}$, and so we define

$$\begin{aligned} \eta_u &= \mu_{u,2} - \mu_{u,1}, & \sigma_{u,i}^2 &= \text{var}(r_{u,i}), \\ \rho_u &= \frac{\text{cov}(r_{u,1}, r_{u,2})}{\sigma_{u,1}\sigma_{u,2}}, & \rho_{i,j} &= \frac{\text{cov}(r_{t,i}, r_{t+1,j})}{\sigma_{t,i}\sigma_{t+1,j}}, \\ \zeta_u^2 &= \sigma_{u,1}^2 + \sigma_{u,2}^2 - 2\rho_u\sigma_{u,1}\sigma_{u,2}, \\ \zeta_{t,t+1} &= \rho_{1,1}\sigma_{t,1}\sigma_{t+1,1} + \rho_{2,2}\sigma_{t,2}\sigma_{t+1,2} - \rho_{1,2}\sigma_{t,1}\sigma_{t+1,2} - \rho_{2,1}\sigma_{t,2}\sigma_{t+1,1}, \\ \varrho_{t,t+1} &= \frac{\zeta_{t,t+1}}{\zeta_t\zeta_{t+1}}, \end{aligned}$$

where $1 \leq i, j \leq 2$ and $u \in \{t, t+1\}$. Note that ς_u^2 is the variance of the spread $r_{u,2} - r_{u,1}$, and $q_{t,t+1}$ is the correlation between $r_{t,2} - r_{t,1}$ and $r_{t+1,2} - r_{t+1,1}$. Next, we compute the terms $\gamma_\tau(x)$ and $Y_\tau(x)$ that appear in the expression (33) for the pdf of the CSM return. For $u, v \in \{t, t+1\}$ and $\tau \in S_2$, we have

$$\begin{aligned}\iota' P_{(1,2)} \mu_u &= -\eta_u = -\iota' P_{(2,1)} \mu_u, \\ D_2 P_{(1,2)} \mu_u &= \eta_u = -D_2 P_{(2,1)} \mu_u, \\ D_2 P_\tau \Sigma_{u,u} P_\tau' D_2' &= \iota' P_\tau \Sigma_{u,u} P_\tau' \iota = \varsigma_u^2, \\ \iota' P_\tau \Sigma_{u,v} P_\tau' D_2' &= -\varsigma_{u,v},\end{aligned}$$

and so

$$\begin{aligned}\gamma_{(1,2)}(x) &= -\eta_{t+1} - \frac{\varsigma_{t,t+1}^2}{\varsigma_t^2} (x - \eta_t), \\ \gamma_{(2,1)}(x) &= \eta_{t+1} - \frac{\varsigma_{t,t+1}^2}{\varsigma_t^2} (x + \eta_t), \\ Y_{(1,2)}(x) &= \frac{\nu + \varsigma_t^{-2} (x - \eta_t)^2}{\nu + 1} \left(\varsigma_{t+1}^2 - \frac{\varsigma_{t,t+1}^2}{\varsigma_t^2} \right) = \frac{\varsigma_{t+1}^2 (1 - q_{t,t+1}^2)}{\nu + 1} \left(\nu + \left(\frac{x - \eta_t}{\varsigma_t} \right)^2 \right), \\ Y_{(2,1)}(x) &= \frac{\nu + \varsigma_t^{-2} (x + \eta_t)^2}{\nu + 1} \left(\varsigma_{t+1}^2 - \frac{\varsigma_{t,t+1}^2}{\varsigma_t^2} \right) = \frac{\varsigma_{t+1}^2 (1 - q_{t,t+1}^2)}{\nu + 1} \left(\nu + \left(\frac{x + \eta_t}{\varsigma_t} \right)^2 \right).\end{aligned}$$

If we define the sign of permutations in S_2 by $\varepsilon((1,2)) = 1$ and $\varepsilon((2,1)) = -1$, then the expressions for $\gamma_\tau(x)$ and $Y_\tau(x)$ can be written succinctly as

$$\gamma_\tau(x) = -\varepsilon(\tau) \eta_{t+1} - q_{t,t+1} \varepsilon(\tau) \left(\frac{x - \varepsilon(\tau) \eta_t}{\varsigma_t} \right), \quad (46)$$

$$Y_\tau(x) = \frac{\varsigma_{t+1}^2 (1 - q_{t,t+1}^2)}{\nu + 1} \left(\nu + \left(\frac{x - \varepsilon(\tau) \eta_t}{\varsigma_t} \right)^2 \right). \quad (47)$$

The next lemma will provide the building blocks for the non-central moments of CSM returns.

Lemma 2. Let $\gamma_\tau(x)$ and $Y_\tau(x)$ be as defined in (46) and (47) respectively, where $\tau \in S_2$. Then, for $\alpha, \beta \in \mathbb{N}$ such that $\alpha + 2\beta < \nu$, we have

$$\begin{aligned}\int_{-\infty}^0 \gamma_\tau^\alpha(x) Y_\tau^\beta(x) t_1(x; \varepsilon(\tau) \eta_t, \varsigma_t^2, \nu) dx \\ = \frac{(-1)^\alpha \varsigma_{t+1}^{2\beta} (1 - \rho_{t,t+1}^2)^\beta}{(\nu + 1)^\beta} \sum_{i=0}^\alpha \sum_{j=0}^\beta \binom{\alpha}{i} \binom{\beta}{j} \varepsilon^{\alpha-i}(\tau) \eta_{t+1}^{\alpha-i} q_{t,t+1}^i \varsigma_{t+1}^i \nu^{\beta-j} \kappa_{i+2j}(\varepsilon(\tau) \eta_t, \varsigma_t^2, \nu),\end{aligned} \quad (48)$$

where $\kappa_m(\eta, \varsigma^2, \nu)$ is as defined in (42).

Proof. Follows from using the binomial formula to expand the powers of $\gamma_\tau(x)$ and $Y_\tau(x)$, and applying the definition of $\kappa_m(\eta, \varsigma^2, \nu)$. \square

For notational convenience, given any $\tau \in S_2$, $\nu \in \mathbb{R}_+$ and $\alpha, \beta \in \mathbb{N}$ such that $\alpha + 2\beta < \nu$, we define

$$\kappa_{\alpha,\beta,\tau}(\eta, \varsigma, \nu) = \int_{-\infty}^0 \gamma_\tau^\alpha(x) Y_\tau^\beta(x) t_1(x; \varepsilon(\tau) \eta, \varsigma^2, \nu) dx, \quad (49)$$

and note that $\kappa_{\alpha,\beta,\tau}(\eta, \varsigma, \nu)$ can be computed explicitly using (48). We now present the non-central moments of the CSM return $r_{1\pm,t+1}$.

Theorem 8. Let $\eta \in \mathbb{R}$, $\varsigma \in \mathbb{R}_+$ and $\nu \in \mathbb{R}_+$. Then, for $m \in \mathbb{N}_+$ such that $m < \nu$, the m -th non-central moment, $\mu_m(r_{1\pm,t+1})$, of $r_{1\pm,t+1}$ is given as follows:

$$\mu_m(r_{1\pm,t+1}) = \sum_{\tau \in S_2} \sum_{\substack{k=0 \\ k \text{ even}}}^m \binom{m}{k} (k-1)!! \frac{\Gamma\left(\frac{1}{2}(\nu-k+1)\right)}{\Gamma\left(\frac{1}{2}(\nu+1)\right)} \left(\frac{\nu+1}{2}\right)^{\frac{k}{2}} \kappa_{m-k, \frac{k}{2}, \tau}(\eta_t, \varsigma_t, \nu), \quad (50)$$

where $\kappa_{\alpha,\beta,\tau}(\eta, \varsigma, \nu)$ is as defined in (49). In particular, the first four non-central moments are

$$\mu_1(r_{1\pm,t+1}) = \sum_{\tau \in S_2} \kappa_{1,0,\tau}(\eta_t, \varsigma_t, \nu), \quad (51)$$

$$\mu_2(r_{1\pm,t+1}) = \sum_{\tau \in S_2} \left(\kappa_{2,0,\tau}(\eta_t, \varsigma_t, \nu) + \frac{(\nu+1)\kappa_{0,1,\tau}(\eta_t, \varsigma_t, \nu)}{\nu-1} \right), \quad (52)$$

$$\mu_3(r_{1\pm,t+1}) = \sum_{\tau \in S_2} \left(\kappa_{3,0,\tau}(\eta_t, \varsigma_t, \nu) + \frac{3(\nu+1)\kappa_{1,1,\tau}(\eta_t, \varsigma_t, \nu)}{\nu-1} \right), \quad (53)$$

$$\mu_4(r_{1\pm,t+1}) = \sum_{\tau \in S_2} \left(\kappa_{4,0,\tau}(\eta_t, \varsigma_t, \nu) + \frac{6(\nu+1)\kappa_{2,1,\tau}(\eta_t, \varsigma_t, \nu)}{\nu-1} + \frac{3(\nu+1)^2\kappa_{0,2,\tau}(\eta_t, \varsigma_t, \nu)}{(\nu-1)(\nu-3)} \right). \quad (54)$$

Proof. Follows from the general expression (38) for the moments of $r_{1\pm,t+1}$ and the definition of $\kappa_{\alpha,\beta,\tau}(\eta, \varsigma, \nu)$. \square

We remark that the non-central moments of $r_{1\pm,t+1}$ given in Theorem 8 are sums indexed by S_2 that consists of two elements, and that each term that appears in these sums can be computed recursively using (43), (48), and (49). Since the right-hand side of (48) consists of a finite number of terms and (43) is equivalent to the explicit expressions (44) or (45) depending on the index m , these moments of $r_{1\pm,t+1}$ can be computed without having to make any simplifying approximations. For example, the first moment is given explicitly by

$$\mu_1(r_{\pm 1,t+1}) = \eta_{t+1} \left(2T_1(0; -\eta_t, \varsigma_t^2, \nu) - 1 \right) + 2\rho_{t,t+1}\varsigma_{t+1} \sqrt{\frac{\nu}{\nu-2}} t_1 \left(0; \eta_t, \frac{\nu\varsigma_t^2}{\nu-2}, \nu \right),$$

which is reassuring since it has the same functional form as the following expression³

$$\mu_1^*(r_{\pm 1,t+1}) = \eta_{t+1} \left(2\Phi\left(\frac{\eta_t}{\varsigma_t}\right) - 1 \right) + 2\rho_{t,t+1}\varsigma_{t+1}\phi\left(\frac{\eta_t}{\varsigma_t}\right)$$

obtained for the normally distributed asset return case in Kwon and Satchell (2018) except for the distribution functions being Student's t rather than normal. The numerical calculations in this paper were performed using code written in C++ that relied on the boost library⁴ to compute the functions t_1 and T_1 .

Returning briefly to the linear factor structure discussed in Section 1, we could consider $\mu_{t,1}$ and $\mu_{t,2}$ to be a linear combinations of factors, which in a Carhart (1997) model context would consist of size, market, value, and momentum. Thus, if we were to go long asset 1 and short asset 2 in our CSM momentum portfolio, we might expect a larger exposure to the momentum factor for asset 1 and

³ Rewritten in the notation of this paper.

⁴ Refer to www.boost.org/doc/libs/1_72_0/boost/math/distributions/students_t.hpp

a smaller exposure for asset 2. We could carry out further detailed analysis to accommodate these features but leave this for further research.

If we denote by $\mu_{1\pm,t+1}$ the mean, $\sigma_{1\pm,t+1}^2$ the variance, $\gamma_{1\pm,t+1}$ the skewness, and $\kappa_{1\pm,t+1}$ the excess kurtosis of the CSM return, then these quantities are easily computed from the non-central moments given in Theorem 8. It should be noted that the quantities corresponding to the odd moments, viz. $\mu_{1\pm,t+1}$ and $\gamma_{1\pm,t+1}$ are approximately odd as functions of $q_{t,t+1}$, and those associated with the even moments, viz. $\sigma_{1\pm,t+1}^2$ and $\kappa_{1\pm,t+1}$, are approximately even as functions of $q_{t,t+1}$. This is because the return from a portfolio formed by taking a long position in the loser and a short position in the winner when $q_{t,t+1} < 0$ would have the same distributional properties as the return from taking the opposite positions when $q_{t,t+1} > 0$.

In the analysis that follows, we assume that the asset returns are stationary in order to reduce the number of parameters. Moreover, we have set $\mu_{t,1} = 6\%$, $\mu_{t,2} = 4\%$, $\sigma_{t,1}^2 = 0.18(\nu - 2)/\nu$ and $\sigma_{t,2}^2 = 0.14(\nu - 2)/\nu$, where the asset variances have been scaled by a factor dependent on ν to ensure that they are independent of the number of degrees of freedom. It should be noted that the cross-sectional correlation, ρ_t , then determines the variance, ζ_t^2 , of the spread, $r_{t,2} - r_{t,1}$.

The mean of the CSM return is shown in Figure 1 as a function of the degrees of freedom ν , and the spread autocorrelation, $q_{t,t+1}$, for $\rho_t = -0.9$, $\rho_t = 0.0$, and $\rho_t = 0.9$. As expected, $\mu_{1\pm,t+1}$ is an increasing function of $q_{t,t+1}$. For $q_{t,t+1} > 0$, the mean decreases slightly with ν , while the opposite is the case when $q_{t,t+1} < 0$. In the region $q_{t,t+1} \approx 0$, the mean is slightly positive. Since this is the region corresponding to small autocorrelations in the underlying asset returns, and the situation most commonly observed in practice, it is reassuring that the small positive CSM returns implied by the model is consistent with the findings reported in the empirical literature. Moreover, we see that the degrees of freedom parameter, ν , that controls the kurtosis of asset returns has very little impact on $r_{1pm,t+1}$. Interpreting small ν as representative of assets from emerging markets, this is consistent with findings from Rouwenhorst (1999) and Bekaert et al. (1997) that, although there is evidence of momentum in emerging markets, it is not significantly different to those observed in developed markets, despite the assets from the respective markets having different distributional properties. The surface flattens out as the cross-sectional correlation increases from -0.9 to 0.9 . Since an increase in ρ_t corresponds to a decrease in the variance, ζ_t^2 , of the spread $r_{t,2} - r_{t,1}$, this behavior is consistent with the positive relationship usually associated with risk and return in finance.

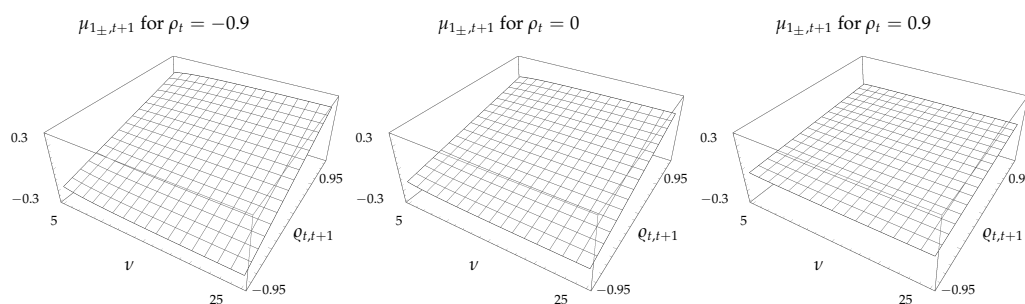


Figure 1. Expected CSM return, $\mu_{1\pm,t+1}$, as a function of ν and $q_{t,t+1}$.

The standard deviation of the CSM return as a function of ν and $q_{t,t+1}$ is shown in Figure 2. Although not clearly evident from the figure, $\sigma_{1\pm,t+1}$ is a decreasing function of ν , and for a fixed value of ν the standard deviation is convex in $q_{t,t+1}$ and takes the maximum value at $q_{t,t+1} = 0$. Finally, since the variance of the spread decreases as ρ_t increases, $\sigma_{1\pm,t+1}$ likewise decreases with increasing ρ_t .

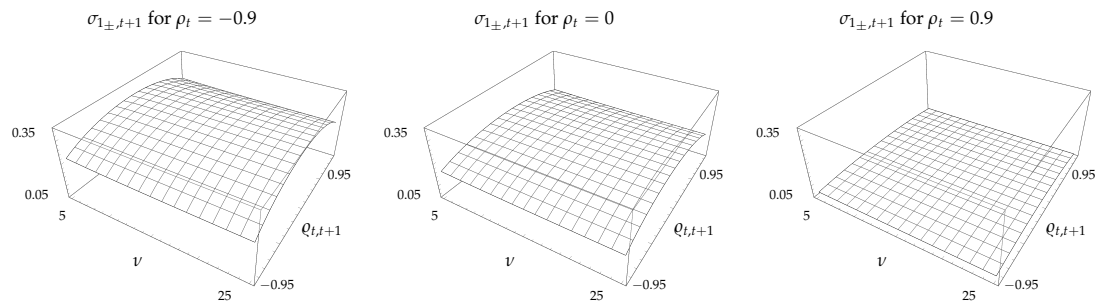


Figure 2. Standard deviation, $\sigma_{1\pm,t+1}$, of CSM return as a function of ν and $q_{t,t+1}$.

The skewness of the CSM return in Figure 3 shows that $\gamma_{1\pm,t+1}$ is negative when $q_{t,t+1} < 0$ and positive otherwise. In fact, although it is not clearly evident from the figure, $\gamma_{1\pm,t+1}$ is negative even for small positive values of $q_{t,t+1}$. As discussed above, the autocorrelations in the asset returns tend to be small in practice, and hence $q_{t,t+1}$ will also be small. The corresponding model implied skewness in the CSM return will then be slightly negative, which is consistent with the observations in the empirical literature.

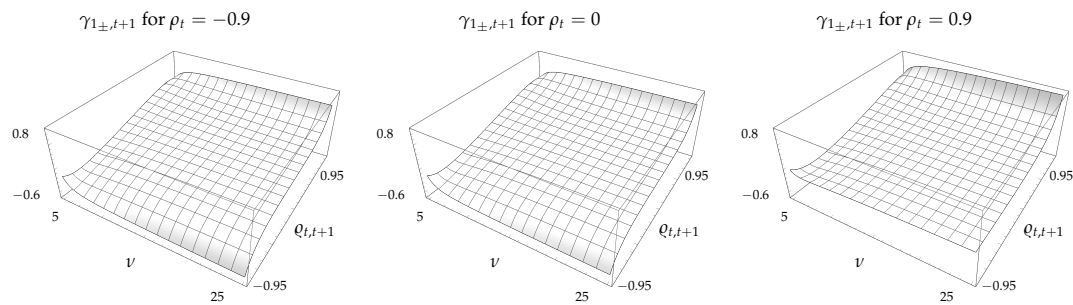


Figure 3. Skewness, $\gamma_{1\pm,t+1}$, of CSM return as a function of ν and $q_{t,t+1}$.

In contrast to the surfaces for other quantities that flatten to a large extent as ρ_t increases from -0.9 to 0.9 , the surface for $\kappa_{1\pm,t+1}$ in Figure 4 remains relatively unchanged. The excess kurtosis of CSM returns is generally positive, in line with the findings reported in the literature, and increases significantly when ν is small and $|q_{t,t+1}|$ is high. It should be noted that $\kappa_{1\pm,t+1}$ is largest when ν is small. Since the deviation of the Student's t distribution from the normal is greatest when ν is small, it follows that the extension considered in this paper will be useful in situations where the observed kurtosis in the CSM returns is higher than the value implied under the assumption of normal asset returns. This would be the case, for example, when considering emerging markets.

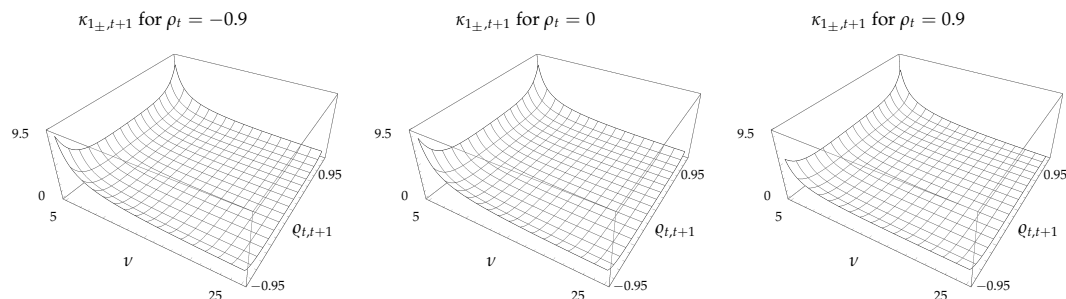


Figure 4. Excess kurtosis, $\kappa_{1\pm,t+1}$, of CSM return as a function of ν and $q_{t,t+1}$.

5. Conclusions

In this paper, the theoretical framework introduced in [Kwon and Satchell \(2018\)](#) was extended to investigate the distributional properties of cross-sectional momentum (CSM) returns under the assumption that the vector of asset returns were multivariate Student's t . The probability density

function and the moments of the CSM returns were derived, and investigated in detail for the special case of two assets.

It was found that, in situations where the assets return, and hence the return spread, autocorrelations are small, and the CSM return has a small positive mean, negative skewness, and excess kurtosis. These are all consistent with the findings reported in the empirical literature. Moreover, the skewness and the kurtosis both become more pronounced as the number of degrees of freedom in the Student's t distribution decreases and the corresponding asset returns become less normal.

In modeling asset returns that deviate significantly from being normal, such as those for emerging markets, the extension to the Student's t considered in this paper would address some of the limitations of assuming normality. Since the Student's t distribution approaches the normal in the limit as the number of degrees of freedom approaches infinity, the extension also provides a framework under which to analyze the implication and the limitations of the assumption of normality in asset returns to CSM returns.

Author Contributions: Conceptualization, O.K.K. and S.S.; methodology, O.K.K. and S.S.; software, O.K.; validation, O.K.K. and S.S.; investigation, O.K.K. and S.S.; writing—original draft preparation, O.K.K. and S.S.; writing—review and editing, O.K.K. and S.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Proof of Theorem 4

From the discussion in Subsection 2.3, see also Roth (2013) Appendix A.4, we have that $\mathbf{X} \sim \text{St}_n(\boldsymbol{\mu}, \Sigma, \nu)$ can be decomposed as

$$\mathbf{X} \stackrel{d}{\sim} \boldsymbol{\mu} + \frac{1}{\sqrt{U}} \mathbf{V},$$

where $U \sim \text{Gam}\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$, $\mathbf{V} \sim \mathcal{N}_n(\mathbf{0}, \Sigma)$, and U and \mathbf{V} are independent. Now, from (9), we obtain

$$\mu_p(\mathbf{X}) = \sum_{P \in \mathcal{P}_m} \prod_{k \in P} \mu_{p_k} \mathbb{E} \left[U^{-\frac{1}{2}|P^c|} \prod_{l \in P^c} V_{p_l} \right],$$

where $P^c = \{1, 2, \dots, m\} \setminus P$ denotes the complement of P , and since U and \mathbf{X} are independent,

$$\mu_p(\mathbf{X}) = \sum_{P \in \mathcal{P}_m} \prod_{k \in P} \mu_{p_k} \mathbb{E} \left[U^{-\frac{1}{2}|P^c|} \right] \mathbb{E} \left[\prod_{l \in P^c} V_{p_l} \right].$$

The second expectation term can be computed using (10), and so it remains to compute the first expectation term. However, from the properties of Gamma distributions, we have

$$\mathbb{E} \left[U^{-\frac{1}{2}|P^c|} \right] = \left(\frac{\nu}{2} \right)^{\frac{1}{2}|P^c|} \frac{\Gamma\left(\frac{\nu}{2} - \frac{1}{2}|P^c|\right)}{\Gamma\left(\frac{\nu}{2}\right)}$$

provided $|P^c| < \nu$, and so substituting into the expression for $\mu_p(\mathbf{X})$ gives (18).

Appendix B. Proof of Theorem 5

For any $k \in \mathbb{N}$, let $\mathbf{1}_k = (1, 1, \dots, 1) \in \mathbb{R}^k$, and define $\iota \in \mathbb{R}^n$ by $\iota = (\mathbf{1}'_{m_+}, \mathbf{0}'_{n-n_+-n_-}, -\mathbf{1}'_{n_-})'$, so that, for any $\tau \in S_n$, we have that $r_{\tau, m_{\pm}, t+1} = \iota' P_{\tau} r_{t+1}$. Then, since

$$\begin{bmatrix} r_{\tau, m_{\pm}, t+1} \\ \mathbf{x}_{\tau, t} \end{bmatrix} = \begin{bmatrix} \iota' P_{\tau} & O_{1 \times n} \\ O_{(n-1) \times n} & D_n P_{\tau} \end{bmatrix} \begin{bmatrix} r_{t+1} \\ \mathbf{r}_t \end{bmatrix},$$

we have that $(r_{\tau, m_{\pm}, t+1}, \mathbf{x}'_{\tau, t})'$ is a linear transformation of $(r'_{t+1}, \mathbf{r}'_t)'$, and so it follows from Roth (2013) Equation (4.1) that

$$\begin{bmatrix} r_{\tau, m_{\pm}, t+1} \\ \mathbf{x}_{\tau, t} \end{bmatrix} \sim \text{St}_n \left(\begin{bmatrix} \iota' P_{\tau} \boldsymbol{\mu}_{t+1} \\ D_n P_{\tau} \mathbf{r}_t \end{bmatrix}, \begin{bmatrix} \iota' P_{\tau} \Sigma_{t+1, t+1} P'_{\tau} \iota & \iota' P_{\tau} \Sigma_{t+1, t} P'_{\tau} D'_n \\ D_n P_{\tau} \Sigma_{t, t+1} P'_{\tau} \iota & D_n P_{\tau} \Sigma_{t, t} P'_{\tau} D'_n \end{bmatrix}, \nu \right).$$

Now, the joint pdf, $f_{r_{\tau, m_{\pm}, t+1}, \mathbf{x}_{\tau, t}}(r, \mathbf{x})$, is given by

$$f_{r_{\tau, m_{\pm}, t+1}, \mathbf{x}_{\tau, t}}(t, \mathbf{x}) = f_{\mathbf{x}_{\tau, t} | r_{\tau, m_{\pm}, t+1}}(\mathbf{x} | r) f_{r_{\tau, m_{\pm}, t+1}}(r).$$

However, from Theorem 3, we have $f_{\mathbf{x}_{\tau, t} | r_{\tau, m_{\pm}, t+1}}(\mathbf{x} | r) = t_{n-1}(\mathbf{x}; \boldsymbol{\lambda}_{\tau}(r), \Lambda_{\tau}(r), \nu + 1)$, and so

$$f_{r_{\tau, m_{\pm}, t+1}, \mathbf{x}_{\tau, t}}(t, \mathbf{x}) = t_{n-1}(\mathbf{x}; \boldsymbol{\lambda}_{\tau}(r), \Lambda_{\tau}(r), \nu + 1) t_1(r; \iota' P_{\tau} \boldsymbol{\mu}_{t+1}, \iota' P_{\tau} \Sigma_{t+1, t+1} P'_{\tau} \iota, \nu).$$

If we define $\mathcal{D}_{\tau} = \{\mathbf{x}_{\tau} \prec \mathbf{0}\}$, then $f_{m_{\pm}, t+1 | \mathcal{D}_{\tau}} = \mathbb{I}_{\mathcal{D}_{\tau}} f_{r_{\tau, m_{\pm}, t+1}}$, and so

$$\begin{aligned} f_{m_{\pm}, t+1 | \mathcal{D}_{\tau}}(r) &= \int_{\mathbf{x} \in \mathbb{R}^{n-1}} \mathbb{I}_{\mathcal{D}_{\tau}}(\mathbf{x}) f_{r_{\tau, m_{\pm}, t+1}, \mathbf{x}_{\tau, t}}(t, \mathbf{x}) d\mathbf{x} \\ &= \int_{\mathbf{x} \prec \mathbf{0}} t_{n-1}(\mathbf{x}; \boldsymbol{\lambda}_{\tau}(r), \Lambda_{\tau}(r), \nu + 1) t_1(r; \iota' P_{\tau} \boldsymbol{\mu}_{t+1}, \iota' P_{\tau} \Sigma_{t+1, t+1} P'_{\tau} \iota, \nu) d\mathbf{x} \\ &= t_1(r; \iota' P_{\tau} \boldsymbol{\mu}_{t+1}, \iota' P_{\tau} \Sigma_{t+1, t+1} P'_{\tau} \iota, \nu) T_{n-1}(\mathbf{0}; \boldsymbol{\lambda}_{\tau, r}, \Lambda_{\tau, r}, \nu + 1), \end{aligned}$$

and summing over $\tau \in S_n$ gives (30). The alternative expression (33) for $f_{m_{\pm}, t+1}(r)$ follows from a similar argument using the decomposition $f_{r_{\tau, m_{\pm}, t+1}, \mathbf{x}_{\tau, t}}(t, \mathbf{x}) = f_{r_{\tau, m_{\pm}, t+1} | \mathbf{x}_{\tau, t}}(r | \mathbf{x}) f_{\mathbf{x}_{\tau, t}}(\mathbf{x})$.

Appendix C. Proof of Theorem 6

In view of the expression for $f_{m_{\pm}, t+1}(r)$ in (33), we need to compute terms of the form

$$\mathcal{I}_{m, \tau} = \int_{\mathbb{R}} \int_{\mathbb{R}^{n-1}} r^m t_1(r; \gamma_{\tau}(\mathbf{x}), Y_{\tau}(\mathbf{x}), \nu + n - 1) t_{n-1}(\mathbf{x}; D_n P_{\tau} \boldsymbol{\mu}_t, D_n P_{\tau} \Sigma_{t, t} P'_{\tau} D'_n, \nu) d\mathbf{x} dr$$

for $\tau \in S_n$. Now, since a one-dimensional Student's t distribution with $\nu - n - 1$ degrees of freedom has finite moments for orders less than the number of degrees of freedom, and by assumption $m < \nu \leq \nu + n - 1$, the m -th moment of $t_1(r; \gamma_{\tau}(\mathbf{x}), Y_{\tau}(\mathbf{x}), \nu + n - 1)$ exists. Next, interchanging the order of integration gives

$$\mathcal{I}_{m, \tau} = \int_{\mathbb{R}^{n-1}} t_{n-1}(\mathbf{x}; D_n P_{\tau} \boldsymbol{\mu}_t, D_n P_{\tau} \Sigma_{t, t} P'_{\tau} D'_n, \nu) \int_{\mathbb{R}} r^m t_1(r; \gamma_{\tau}(\mathbf{x}), Y_{\tau}(\mathbf{x}), \nu + n - 1) dr d\mathbf{x},$$

and applying Corollary 2 to the inner integral gives

$$\mathcal{I}_{m,\tau} = \sum_{\substack{k=0 \\ k \text{ even}}}^m \binom{m}{k} (k-1)!! \frac{\Gamma\left(\frac{1}{2}(\nu+n-1-k)\right)}{\Gamma\left(\frac{1}{2}(\nu+n-1)\right)} \left(\frac{\nu+n-1}{2}\right)^{\frac{k}{2}} \int_{\mathbb{R}_{-}^{n-1}} \gamma_{\tau}^{m-k}(\mathbf{x}) Y_{\tau}^{\frac{k}{2}}(\mathbf{x}) t_{n-1}(\mathbf{x}; D_n P_{\tau} \boldsymbol{\mu}_t, D_n P_{\tau} \Sigma_{t,t} P'_{\tau} D'_n, \nu) d\mathbf{x}.$$

Since $\gamma_{\tau}(\mathbf{x})$ and $Y_{\tau}(\mathbf{x})$ are of orders 1 and 2 respectively in \mathbf{x} , we have $\gamma_{\tau}^{m-k}(\mathbf{x}) Y_{\tau}^{\frac{k}{2}}(\mathbf{x})$ is of order m in \mathbf{x} , and since $t_{n-1}(\mathbf{x}; D_n P_{\tau} \boldsymbol{\mu}_t, D_n P_{\tau} \Sigma_{t,t} P'_{\tau} D'_n, \nu)$ has finite moments of order up to ν by Theorem 4, it follows that the integral in $\mathcal{I}_{m,\tau}$ is well-defined. Summing the $\mathcal{I}_{m,\tau}$ over $\tau \in S_n$ gives (38).

Appendix D. Proof of Lemma 1

Setting the scale parameter to $\varsigma \sqrt{\nu/(\nu-2)}$ and the degrees of freedom to $\nu-2$ in (40) gives

$$t_1\left(x; \eta, \varsigma^2 \nu / (\nu-2), \nu-2\right) = c\left(\eta, \varsigma^2 \nu / (\nu-2), \nu-2\right) \left(1 + \frac{1}{\nu} \left(\frac{x-\eta}{\varsigma}\right)^2\right)^{-\frac{1}{2}(\nu-1)},$$

and differentiating both sides with respect to x gives

$$\begin{aligned} \frac{\partial t_1(x; \eta, \varsigma^2 \nu / (\nu-2), \nu-2)}{\partial x} &= -c\left(\eta, \frac{\varsigma^2 \nu}{\nu-2}, \nu-2\right) \frac{(\nu-1)}{\nu \varsigma} \left(\frac{x-\eta}{\varsigma}\right) \left(1 + \frac{1}{\nu} \left(\frac{x-\eta}{\varsigma}\right)^2\right)^{-\frac{1}{2}(\nu+1)} \\ &= -\frac{c(\eta, \varsigma^2 \nu / (\nu-2), \nu-2)(\nu-1)}{c(\eta, \varsigma^2, \nu) \nu \varsigma} \left(\frac{x-\eta}{\varsigma}\right) t_1(x; \eta, \varsigma^2, \nu). \end{aligned} \quad (\text{A1})$$

Now, using the definition of $c(\eta, \varsigma, \nu)$, we have

$$\frac{c(\eta, \varsigma^2 \nu / (\nu-2), \nu-2)(\nu-1)}{c(\eta, \varsigma^2, \nu) \nu \varsigma} = \frac{\Gamma\left(\frac{\nu-1}{2}\right)}{\varsigma \sqrt{(\nu-2)\pi} \Gamma\left(\frac{\nu-2}{2}\right)} \cdot \frac{\varsigma \sqrt{\nu\pi} \Gamma\left(\frac{\nu}{2}\right)}{\Gamma\left(\frac{\nu+1}{2}\right)} \cdot \frac{(\nu-1)}{\nu \varsigma},$$

and using the property $\Gamma(z+1) = z\Gamma(z)$, we obtain

$$\frac{c(\eta, \varsigma^2 \nu / (\nu-2), \nu-2)(\nu-1)}{c(\eta, \varsigma^2, \nu) \nu \varsigma} = \sqrt{\frac{\nu}{\nu-2}} \cdot \frac{\frac{1}{2}(\nu-2)}{\frac{1}{2}(\nu-1)} \cdot \frac{(\nu-1)}{\nu \varsigma} = \frac{1}{\varsigma} \sqrt{\frac{\nu-2}{\nu}}.$$

Substituting back into (A1) and rearranging gives (41) for $m=1$, and the general case follows from multiplying both sides of the identity for the $m=1$ case by powers of $(x-\eta)/\varsigma$.

References

- Arellano-Valle, Reinaldo B., and Adelchi Azzalini. 2006. On the Unification of Families of Skew-normal Distributions. *Scandinavian Journal of Statistics* 33: 561–74.
- Arellano-Valle, Reinaldo B., and Marc G. Genton. 2010. Multivariate unified skew-elliptical distributions. *Chilean Journal of Statistics* 1: 17–33.
- Asness, Clifford S. 1994. Variables that Explain Stock Returns. Ph.D. dissertation, University of Chicago, Chicago, IL, USA.
- Asness, Clifford S., John M. Liew, and Ross L. Stevens. 1997. Parallels between the cross-sectional predictability of stock and country returns. *Journal of Portfolio Management* 23: 79–87.
- Asness, Clifford S., Tobias J. Moskowitz, and Lasse H. Pedersen. 2013. Value and momentum everywhere. *The Journal of Finance* 58: 929–85.
- Azzalini, Adelchi, and Alessandra Dalla Valle. 1985. The multivariate skew-normal distribution. *Biometrika* 83: 715–26.
- Bekaert, Geert, Claude B. Erb, Campbell R. Harvey, and Tadas E. Viskanta. 1997. What Matters for Emerging Equity Market Investments. *Emerging Markets Quarterly* 1: 17–46.
- Carhart, Mark M. 1997. On Persistence in Mutual Fund Performance. *Journal of Finance* 52: 57–82.
- Chan, Kalok, Allaudeen Hameed, and Wilson Tong. 2000. Profitability of Momentum Strategies in the International Equity Markets. *Journal of Financial and Quantitative Analysis* 35: 153–72.
- Daniel, Kent, and Tobias J. Moskowitz. 2016. Momentum Crashes. *Journal of Financial Economics* 122: 221–47.
- Erb, Claude B., and Campbell R. Harvey. 2006. The strategic and tactical value of commodity futures. *Financial Analysts Journal* 62: 69–97.
- Fama, Eugene, and Kenneth French. 1992. The Cross-Section of Expected Stock Returns. *The Journal of Finance* 47: 427–65.
- Hameed, Allaudeen, and Kusnadi Yuanto. 2002. Momentum Strategies: Evidence from the Pacific Basin Stock Markets. *Journal of Financial Research* 25: 383–97.
- Hansen, Christian, James B. McDonald, and Whitney K. Newey. 2010. Instrumental Variables Estimation with Flexible Distributions. *Journal of Business and Economic Statistics* 28: 13–25.
- Israel, Ronen, and Tobias J. Moskowitz. 2013. The Role of Shorting, Firm Size, and Time on Market Anomalies. *Journal of Financial Economics* 108: 275–301.
- Jamalizadeh, Ahad, and Narayanaswamy Balakrishnan. 2012. Concomitants of order statistics from multivariate elliptical distributions. *Journal of Statistical Planning and Inference* 142: 397–409.
- Jegadeesh, Narasimhan, and Sheridan Titman. 1993. Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance* 48: 65–91.
- Jegadeesh, Narasimhan, and Sheridan Titman. 2001. Profitability and Momentum Strategies: An Evaluation of Alternative Explanations. *Journal of Finance* 56: 699–720.
- Kan, Raymond. 2008. From moments of sum to moments of product. *Journal of Multivariate Analysis* 99: 542–54.
- Kirkby, Justin L., Dang Nguyen, and Duy Nguyen. 2019. Moments of Student's t-distribution: A Unified Approach. *arXiv* 2019, arXiv:1912.01607.
- Kwon, Oh Kang, and Stephen Satchell. 2018. The distribution of cross sectional momentum returns. *Journal of Economic Dynamics and Control* 94: 225–41.
- Lewellen, Jonathan. 2002. Momentum and autocorrelation in Stock Returns *Review of Financial Studies* 15: 533–63.
- Lo, Andrew, and Craig MacKinlay. 1990. When are Contrarian Profits due to Stock Market Overreaction? *Review of Financial Studies* 3: 175–205.
- Menkhoff, Lucas, Lucio Sarno, Maik Schmeling, and Andreas Schrimpf. 2012. Currency momentum strategies. *Journal of Financial Economics* 106: 660–84.
- Moskowitz, Tobias J., Yao Hua Ooi, and Lasse H. Pedersen. 2012. Time Series Momentum. *Journal of Financial Economics* 104: 228–50.
- Muirhead, Robb J. 1982. *Aspects of Multivariate Statistical Theory*. New Jersey: John Wiley & Sons.
- Okunev, John U., and Derek White. 2003. Do momentum-based strategies still work in foreign currency markets? *Journal of Financial and Quantitative Analysis* 38: 425–47.
- Richards, Anthony J. 1997. Winner-Loser Reversals in National Stock Market Indices: Can They be Explained?. *Journal of Finance* 52: 2129–44.

- Roth, Michael. 2013. *On the Multivariate t-Distribution*. Technical Report. Linköping: Department of Electrical Engineering, Linköpings Universitet.
- Rotman, Joseph J. 1995. *An Introduction to the Theory of Groups*. New York: Springer-Verlag.
- Rouwenhorst, Geert K. 1998. International Momentum Strategies. *Journal of Finance* 53: 267–84.
- Rouwenhorst, Geert K. 1999. Local Return Factors and Turnover in Emerging Stock Markets. *Journal of Finance* 54: 1439–64.
- Sefton, James, and Alan Scowcroft. 2004. *A Decomposition of Portfolio Momentum Returns*. Discussion Paper TBS/DP04/9. London: Tanaka Business School, Imperial College London.
- Theodossiou, Panayiotis. 1998. Financial Data and the Skewed Generalized T Distribution. *Management Science* 44: 1650–61.
- Withers, Christopher S. 1985. The Moments of the Multivariate Normal. *Bulletin of Australian Mathematics Society* 32: 103–7.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).